

Protocol for Calibrating and Scoring ELA Tasks

Rhode Island Skills Commission

Explanation and Considerations for Use

The Rhode Island Skills Commission and its network schools developed this document. It represents one approach; your school may choose to adopt it or may prefer to explore other approaches to assessing common tasks.

This protocol is used to familiarize all scorers with the elements required to judge the quality of a student's work on a given common task. It entails examination of the standards addressed in the task, the clarity of teacher and student directions, the prompt, and the rubric. Examination and discussion of at least two different samples of student work is recommended.

This tool was created and/or compiled by The Rhode Island Department of Education and The Education Alliance at Brown University, with the generous support of the Bill & Melinda Gates Foundation.

<http://www.ride.ri.gov/highschoolreform/dslat/>
October, 2005



Protocol for Calibrating and Scoring ELA Tasks: RI Skills Commission

Introduction

Calibrating tasks is both an art and a science. Because the calibration process uses discussion to arrive at understanding, it is an art—the art of focused conversation. And because calibration aims at developing a precise understanding of how a rubric measures student work, it is a science—the science of measurement.

In calibration, a group of scorers all look at the same piece of work, score the work using a rubric, and then discuss how they scored the work. Disagreements within the group regarding the scores lead to focused discussions about scoring. These discussions typically include two major topics: the weight and worth of evidence from the work, and what the rubric means. As the discussion continues, problems and solutions are recorded as “calibration notes” for future scorers. Particularly clear examples of work at different levels are also kept for future scorers as benchmarks.

Preparation

Someone who has previous experience in the calibration and scoring process should lead the calibration session. Your school should have someone like this on staff and, if not, should arrange to have an experienced outside leader facilitate the calibration process. The best way for your school to develop an experienced leader in calibration is to have someone attend the Skills Commission’s central scoring sessions several times.

Calibration requires a *scoring packet*. A scoring packet consists of copies of 3 to 5 pieces of the work that is going to be scored. There should be as many copies of each piece of work as there are members of the calibration team. Each piece of work should have a rubric attached to it. The leader of the calibration process should also come equipped to take notes on the calibration process or designate a member of the group to do this.

The calibration group should be seated in such a way that every member can talk to, and hear, every other member of the group.

The Process

The leader of the calibration process distributes the first piece of work to every member of the group (every member gets a copy of the same piece of work). The group reads the work and uses the rubric to score the work. This is done by highlighting the indicators within each box that best describe the work. When all members of the group have finished scoring the piece of work, the leader begins the calibration discussion.

First, the leader gets a count of the way the group scored the work *as a whole* (the overall score). The leader records how many people scored the work as “exceeding the standard”, “meeting the standard”, “nearly meeting the standard”, “below standard” and “little or no evidence of the standard”. This gives the leader and the group a sense of how much consensus there is to begin with.

Next, the leader begins with the first row on the rubric and repeats the question for that row. The leader then begins the calibration discussion by asking someone from one of the extreme scores to explain why they gave the score they did. This explanation should refer as carefully as possible to evidence from the text that is relevant to the rubric. For example, if the scorer thinks the work is below standard for “usage and grammar”, the scorer would point to the instances of usage and grammar in the work sample that led to assigning this score.

The calibration leader would then ask someone who had given a score at the opposite extreme to explain why he/she gave the score they did. This person could also comment on the evidence the previous person used. This begins a more general discussion about the particular points that are used to assign scores. The leader should check for changes in the group’s judgment from time to time by polling how many scores are at the different levels as points are resolved.

If the group is not making progress towards resolution, the leader should ask the groups to identify the issues that still prevent consensus and then the group should discuss these issues using the same process described above. When the group arrives at resolution (or as close to resolution as possible), it moves on to the next row in the rubric until it has covered the whole rubric. Once this is done, the leader checks for the overall score again and usually, by this time, the group has arrived at clear agreement on the overall score.

The leader ensures that agreements over problematic issues, wording in the rubric, and other issues are recorded for future calibration sessions and for editing the task and rubric at a future time.

The calibration discussion is then repeated for a second piece of student work. It is preferable to avoid pieces that are obviously below standard and lower for calibration discussions. After calibrating a second piece of work, the leader makes a decision about how many more pieces of work the group needs to calibrate before it has a clear sense of how to apply the rubric. Often the group achieves this after two pieces of work, but it may take as many as five pieces of work for complex tasks.

Scoring

Paired Scoring: Calibration can be carried into the scoring process by using a “paired scoring” process. In paired scoring, two scorers work together by scoring the same pieces of work. After both scorers have finished, they compare their scores for over all agreement. If they do not have overall agreement, they look at the way they scored the

rows in the rubric to find those places where they disagreed. They then discuss these areas using evidence from the work in an attempt to resolve their disagreements. If this is possible, one of the two scorers changes his/her score and the pair moves on to the next piece of work. If it is not possible, they pair consults a third, expert, scorer who reads the work and explains how s/he would score the work. The work then receives the three scores plus the final overall score and the paired scorers move on.

Open Table Scoring: A group of calibrators score a set of tasks and consistently, or on a sampling basis, select a task that has been scored by someone else to score a second time. The second scoring should be done without looking at the way the first scorer filled in the rubric and if the second overall score agrees with the first, the scorer move on. If the two scorers do not agree, they follow the procedure described above, including adjudication by a third expert scorer if necessary.